

On the origin of Hilbert Geometry

Marc Troyanov - EPFL

July 15, 2014

Abstract

In this brief essay we succinctly comment on the historical origin of Hilbert geometry. In particular, we give a summary of the letter in which David Hilbert informs his friend and colleague Felix Klein about his discovery of this geometry. The present paper is to appear in the Handbook of Hilbert geometry, (ed. A. Papadopoulos and M. Troyanov), European Mathematical Society, Zürich, 2014.

Hilbert geometry, which is the main topic of this handbook, was born twelve decades ago, in the summer of 1894. Its birth certificate is a letter written by David Hilbert to Felix Klein, on August 14, 1894, sent from the village of Kleinteich near the city of Rauschen (today called Svetlogorsk), at the Baltic sea. The place was a fashionable vacation resort, situated some forty kilometers north of Hilbert's hometown Königsberg (now Kaliningrad), where he was a Privatdozent¹. The mathematical part of this letter was published one year later in the *Mathematische Annalen*² under the title *Ueber die gerade Linie als kürzeste Verbindung zweier Punkte* (On the straight line as shortest connection between two points) [6].

Before discussing the content of this letter, let us say a few words on the general situation of geometry in the XIXth century. At the dawn of the century, and despite numerous criticisms concerning the axioms, Euclid's *Elements* were still considered as a model of mathematical rigor and an exposition of the "true" geometry. The 1820s saw on one hand the creation of non-Euclidean geometry by Gauss, Bolyai and Lobachevsky, and on the other hand, a revival and a rapid development of projective geometry by Poncelet, Gergonne, Steiner, Möbius, Plücker and von Staudt. During this period, a controversy developed among supporters of the synthetic versus the analytic methods and techniques in the subject. In this context, the *analytic methods* are based on coordinates and algebraic relations for the description and the analysis of geometric figures, while the *synthetic methods* are based on the incidence relations between points, lines, planes and other loci. The key to unify both sides of the controversy came from

¹Hilbert was appointed professor at Göttingen one year later.

²The chief editor of this journal was Klein himself.

the work of von Staudt [18, 19] who proved that numerical coordinates can be assigned to points on a line in a projective space through the use of purely synthetic methods and without using distances. Von Staudt's theorem implies that the cross ratio of four aligned points can be defined synthetically, and thus it is invariant under collineations (that is, transformations preserving lines and incidence).

Recall that the cross ratio of four points X, Y, Z, T is analytically defined as

$$[X, Y, Z, T] = \frac{(x - z)}{(x - t)} \cdot \frac{(y - t)}{(y - z)},$$

where $x, y, z, t \in \mathbb{R} \cup \{\infty\}$ are coordinates on the given line, corresponding to the points X, Y, Z, T . The synthetic definition of the cross ratio is more elaborate and it is based on an iteration of the construction of the harmonic conjugate of a point with respect to a given pair of points (the harmonic conjugate corresponds to a cross ratio equal to -1 and can easily be constructed synthetically using complete quadrangles), together with ordering and density arguments. In fact, von Staudt was not able to fully prove his theorem since at his time the topological structure of the real numbers, in particular the completeness axiom, had yet to be clarified. This was done later on in Dedekind's *Was sind und was sollen die Zahlen?* (1888). A full exposition of von Staudt's construction is given in chapter VII of Veblen and Young's book [16], see also [15, chap. VII]. We refer to [10, 17] for additional comments on von Staudt's work.

By the middle of the XIXth century, it was seen as a natural problem to try to define metric notions from projective ones, thus reversing the analytic way of seeing geometry. Von Staudt's theorem allowed the use of the cross ratio in this adventure. The first step in this direction is usually attributed to Laguerre, who proved in 1853 that the Euclidean angle φ between two lines ℓ, m through the origin in a plane is given by

$$\varphi = \left| \frac{1}{2i} \log[X, Y, U, V] \right|, \quad (1)$$

where $X \in \ell$, $Y \in m$ and U, V are points representing the “isotropic lines” of the plane³. These are lines which are “orthogonal to themselves”, meaning that if $U = (u_1, u_2)$ and $V = (v_1, v_2)$, then

$$u_1^2 + u_2^2 = v_1^2 + v_2^2 = 0,$$

or, equivalently, $u_2 = iu_1$ and $v_2 = -iv_1$. To explain Laguerre's formula, one may assume $x_1 = y_1 = u_1 = v_1 = 1$; then $x_2 = \tan(\alpha)$, $y_2 = \tan(\beta)$, $u_2 = i$ and $v_2 = -i$. We then have

$$[X, Y, U, V] = \frac{(\tan(\alpha) - i)}{(\tan(\alpha) + i)} \cdot \frac{(\tan(\beta) + i)}{(\tan(\beta) - i)} = e^{2i(\alpha - \beta)}.$$

³In fact the formula in Laguerre's paper does not explicitly involve the cross ratio, the present formulation seems to be due to F. Klein, see e.g. [9, p. 158].

Klein in [8], inspired by a formula published by Cayley⁴ in [3], observed later on that Equation (1) also gives a projective definition of the distance in elliptic geometry, that is, the metric on \mathbb{RP}^n for which the standard projection $\mathbb{S}^n \rightarrow \mathbb{RP}^n$ is a local isometry. Here X and Y are arbitrary distinct points in \mathbb{RP}^n and U and V are points in \mathbb{CP}^n which are aligned with X and Y and belong to the quadric

$$\sum_{i=0}^n u_i^2 = 0. \quad (2)$$

Klein also considered similar geometries where the quadric (2) is replaced by an arbitrary quadric, which, following Cayley's terminology, he called *the absolute*. In 1871, Klein realized that choosing as the absolute the standard cone $\sum_{i=1}^n u_i^2 - u_0^2 = 0$, the metric restricted to the interior of the cone is a model of Lobatchevsky's geometry. See [2] for a historical discussion.

In 1888, M. Pasch proposed a new axiomatic foundation of geometry based on the primitive notion of segment (or betweenness) rather than the projective notion of lines. Pasch's geometry is often called *ordered geometry* and it greatly influenced Hilbert.

Let us now return to Hilbert's letter. The letter starts with a discussion of how geometry should be founded on three "elements" (in the sense of primitive notions) named points, lines and planes (Hilbert's aim is a discussion of three-dimensional geometry). These elements should satisfy three groups of axioms, which he only shortly discusses.

I. The first groups are axioms of incidence describing the mutual relations among points, lines and planes. These axioms state that two distinct points determine a unique line, that three non aligned points determine a unique plane, that a plane containing two distinct points contains the full line determined by those points, and that two planes cannot meet at only one point. Furthermore, every line contains at least two points, every plane contains at least three non aligned points and the space contains at least four non coplanar points.

II. The second groups of axioms deals with ordering points on a line and they have been proposed by M. Pasch in [12]. They state: between two distinct points A and B on a line, there exists at least a third point C , and given three points on a line, one and only one of them lies between the two others. Given two distinct points A and B , there is another point C on the same line such that B is between A and C . Given four points on a line, we can order them as A_1, A_2, A_3, A_4 in such a way that if $h < i < k$, then A_i lies between A_h and A_k . If a plane α contains a line a , then the plane is separated in two half-planes such that A and B belong to the same half-plane if and only if there is no point of the line a between A and B .

III. The axiom of continuity. Let A_i be an infinite sequence of points on a line a . If there exists a point B such that A_i lies between A_k and B as soon as $k \leq i$, then there exists a point C on the line a such that A_i lies between A_1 and C

⁴Cayley's formula also did not involve the cross ratio; it is written in terms of homogenous coordinates.

for any $i > 1$ and any point C' with the same property lies between C and B . (We recognize an axiom stating that any decreasing sequence on a line that is bounded below has a unique greatest lower bound).

Hilbert then claims that the von Staudt theory of harmonic conjugates can be fully developed on this axiomatic basis (in a “similar way” as in Lindemann’s book, *Vorlesungen über Geometrie*). It therefore follows that coordinates can be introduced. More precisely, to each point, one can associate three real numbers x, y, z and each plane corresponds to a linear relation among the coordinates. Furthermore, if one interprets x, y, z as rectangular coordinates in the usual Euclidean space, then the points in our initial space correspond to a certain convex⁵ body in Euclidean space. Conversely, the points in an arbitrary convex domain represent points in our initial space: *our initial space is build from the points in a convex region of the Euclidean space.*

The axiomatic framework set here by Hilbert is interesting in several aspects. The first group of axioms belongs to what is now called *incidence geometry*. Observe that Hilbert formulates them for a three-dimensional space – at that time it was unusual to discuss geometry in an arbitrary dimension n – but note also that this restriction reduces the complexity of the theory, because it is not so simple to define and discuss a general notion of dimension in pure incidence geometry. Also Desargues theorem (which plays a key role in the introduction of coordinates) need not be added as an axiom since it can be proved in this framework.

The second group of axioms describes *ordered geometry*, see [15] and [14] for readable accounts of this topics and [13] for an impressive historical compilation of the literature since Pasch’s original work. Notice that the notion of line (and hence plane) can be defined on the sole base of the primitive notion of segment, but Hilbert found it convenient to have lines and planes as primitive notions also. Again, this reduces the complexity of the theory (and furthermore it is in Euclid’s spirit).

In a more contemporary language, what Hilbert is claiming here is that *an abstract space M consisting of points, together with a structure made of lines, planes and the betweenness relation (for points on a line) satisfying the said axioms is isomorphic to a convex region Ω in \mathbb{R}^3 with the usual notions of lines and planes and the condition that C lies between A and B if and only if $C \in [A, B]$.* It is not clear whether the axiomatics proposed in Hilbert’s letter is sufficient to rigorously prove this result, and the question is also not addressed in his *Grundlagen* [7]. However it is indeed the case that a synthetic axiomatic characterization of convex domains in \mathbb{R}^n can be based on ordered geometry. A theorem of this type is given in W.A. Coppel’s book, who calls it *the fundamental theorem of ordered geometry*, see [4, p. 173].

At this point of his letter, Hilbert stresses that arbitrary convex bodies also appear in Minkowski’s work on number theory. He then proceeds to define a

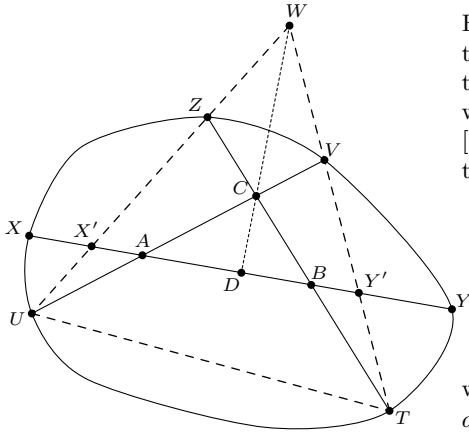
⁵Hilbert, following Minkowski, uses the expression *nirgend concaven Körper*, which means *nowhere concave*. In other words, the domain is convex, but not necessarily strictly convex.

notion of length for the segment AB in his general geometry, which he assumes is now realized as the interior of a convex region in \mathbb{R}^3 . He follows Klein's construction and defines the distance between two distinct points A and B to be⁶

$$d(A, B) = \log[X, Y, B, A] = \log \left(\frac{\overline{YA}}{\overline{YB}} \cdot \frac{\overline{XB}}{\overline{XA}} \right), \quad (3)$$

where X, Y are the two points on the boundary of the convex domain aligned with A and B in the order X, A, B, Y . Observe that $d(A, B) > 0$ since $\frac{YA}{AB}, \frac{XB}{XA} > 1$ (if $A = B$ it is understood that $d(A, B) = 0$).

Hilbert observes that the distance (3) depends on the given convex domain: *if X approaches A and Y approaches B , then the distance $d(A, B)$ increases.* In other words, a smaller convex domain gives rise to larger Hilbert distances. Hilbert then proceeds to give a proof of the triangle inequality.



Here is the argument with Hilbert's notation: From the invariance of the cross ratio with respect to the perspective at W , we have $[U, V, C, A] = [X', Y', D, A]$ and $[Z, T, B, C] = [X', Y', B, D]$. Multiplying these identities gives

$$\begin{aligned} & [U, V, C, A] \cdot [Z, T, B, C] \\ &= [X', Y', D, A] \cdot [X', Y', B, D] \\ &= [X', Y', B, A] \\ &\geq [X, Y, B, A], \end{aligned}$$

which is equivalent to $d(A, C) + d(C, B) \geq d(A, B)$.

Hilbert also observes that the triangle inequality degenerates to an equality when the three points are aligned with C between A and B . Furthermore, he discusses necessary and sufficient conditions for the equality in the triangle inequality. He shows that a non-degenerate triangle exists for which the sum of two sides is equal to the third if and only if there exists a plane which meets the boundary of the convex domain on two segments not on the same line (these would be the segments $[V, T]$ and $[Z, U]$ on the figure). He quotes, as a particularly interesting example, the case where the convex domain is bounded by a tetrahedron.

As a final word, Hilbert points out that he always assumed the given convex body to be bounded, and this hypothesis implies that his geometry does not satisfy the Euclidean parallel postulate.

Let us conclude with two remarks. We first mention that Hilbert came back at least at two occasions on the mathematical subjects discussed in his letter. First

⁶the usual convention now is to divide this quantity by 2.

in his book on the foundation of geometry [7], the book in which he systematically develops the axiomatics of Euclidean 3-space, starting with the same axioms of incidence, order and continuity, to which he adds some axioms on congruence and the parallel postulate. The second occasion is the 1900 International Congress in Paris where he states the famous Hilbert problems. Problem IV concerns *the construction and systematic treatment of the geometries* in convex domains for which straight lines are the shortest, that is, the case where the triangle inequality is an equality in the case of three aligned points. Minkowski and Hilbert geometries are natural examples of such geometries. We refer to the discussion in the last chapter of this Handbook [11].

The second and last remark is about how Hilbert understood the word *geometry*. For Hilbert, a geometry was not conceived as an abstract metric space⁷ satisfying some specific axioms, but rather, as we saw, a geometry is a system made of points, lines and planes subject to a system of consistent interrelations which are accepted as axioms. A notion of distance between points in a given geometry has then to be constructed from the given data and axioms and the properties of the distance, including the triangle inequality, is then a theorem that needs a proof rather than an axiom or a part of the initial definition. This is also in the spirit of Euclid's *Element*, where the triangle inequality is proved in Book 1, Proposition 20.

The subject of *metric geometry* has been developed since the 1920s by a long list of mathematicians including Hausdorff, Menger, Blumenthal, Urysohn, Birkhoff, Busemann, Alexandrov, and others. Since the work of Gromov in the 1980s, metric geometry is seen as a topic of major importance in geometry.

Acknowledgement. The author thanks A. Papadopoulos and K.-D. Semmler for carefully reading the manuscript and providing useful comments.

References

- [1] H. Busemann, *The geometry of geodesics*, Academic Press (1955), reprinted by Dover in 2005.
- [2] N. A'Campo and A. Papadopoulos, in: Sophus Lie and Felix Klein: The Erlangen program and its impact in mathematics and physics (ed. L. Ji and A. Papadopoulos), European Mathematical Society, Zürich, 2014.
- [3] A. Cayley, A sixth Memoir upon Quantics, Philosophical Transactions of the Royal Society of London, Vol. 149 (1859), pp. 61-90.
- [4] W. A. Coppel, *Foundations of convex geometry* Australian Mathematical Society lecture series, 12 Cambridge University Press, (1998).
- [5] M. Fréchet, Sur quelques points du calcul fonctionnel, Rendiconti del Circolo Matematico di Palermo 22 (1906), pp 1-72.

⁷The notion of abstract metric space first appeared in the 1906 thesis of M. Fréchet [5].

- [6] D. Hilbert, Über die gerade Linie als kürzeste Verbindung zweier Punkte, *Math. Ann.*, 46 (1895), 91-96.
- [7] D. Hilbert, Grundlagen der Geometrie, B.G. Teubner, Leipzig, 1899 (various editions and translations).
- [8] F. Klein, Über die sogenannte Nicht-Euklidische Geometrie Mathematische Annalen IV (1871), 573–625.
- [9] F. Klein, *Vorlesungen über höhere Geometrie*, Dritte Auflage. Springer Verlag 1926.
- [10] P. Nabonnand, La théorie des “Würfe” de von Staudt—une irruption de l’algèbre dans la géométrie pure. *Arch. Hist. Exact Sci.* 62 (2008), no. 3, 201–242.
- [11] A. Papadopoulos, On Hilbert’s Fourth Problem, Handbook of Hilbert geometry, (ed. A. Papadopoulos and M. Troyanov), European Mathematical Society, Zürich, 2014.
- [12] M. Pasch, *Vorlesungen über neuere Geometrie* (Leipzig, 1882).
- [13] V. Pambuccian, The axiomatics of ordered geometry: I. Ordered incidence spaces, *Expo. Math.* 29, no. 1, 24–66. (2011).
- [14] W. Prenowitz, and M. Jordan, *Basic concepts of geometry*. Blaisdell Publishing Company, a division of Ginn and Company. (1965).
- [15] G. Robinson, *The Foundations of Geometry* Mathematical Expositions, no. 1. University of Toronto Press, Toronto, Ont., (1940).
- [16] O. Veblen and J.W. Young, *Projective geometry*. Vol. 1. Ginn and Co. (1910).
- [17] J.-D. Voelke, Le théorème fondamental de la géométrie projective: évolution de sa preuve entre 1847 et 1900. *Arch. Hist. Exact Sci.* 62 (2008), no. 3, 243–296.
- [18] K. von Staudt *Geometrie der Lage*, Nürnberg 1847.
- [19] K. von Staudt *Beitrage zur Geometrie der Lage*, Nürnberg 1857.

Marc Troyanov, Section de Mathématiques,
 École Polytechnique Fédérale de Lausanne,
 SMA–Station 8,
 CH-1015 Lausanne - Switzerland
 marc.troyanov@epfl.ch